文化在人机交互中的角色：人类共享和文化差异

胡晓檬 朱昕语 李南瑛 黄歆睿 杜青青 刘若婷 彭栗雨涵 彭凯平

（中国人民大学心理学系，北京，100872）

摘要：随着人工智能（AI）技术的迅速发展和广泛应用，其对社会文化的影响引发了广泛关注。本文通过文化心理学的视角，探讨了文化对 AI 理解和应用的影响，突出了在 AI 开发过程中考虑文化多样性的重要性。文章分析了文化差异如何塑造 AI 技术的接受度和用户体验和不同文化背景下对 AI 的共同的价值观和态度。此外，还讨论了文化背景在 AI 技术的伦理监管和实际应用中的作用。在此基础上，本文提出未来研究方向，强调了考虑 AI 研究中文化特征的重要性，并探索了样本多样性和多种数据统计方法对研究结果的潜在影响。为了全面理解 AI 技术与文化间的相互作用，未来研究需致力于开发适用于不同文化和社会背景的有效评估工具，以促进 AI 技术的全球化发展和应用。

关键词：人工智能，文化心理学，文化差异，文化共享，AI 伦理

1 引言

随着人工智能 (AI)技术的不断发展和广泛应用，其在改变我们生活和工作方式的同时，也引起了越来越多学者对其社会文化影响的关注。AI 涵盖了从简单算法到复杂的机器学习系统的广泛技术，这些技术用以执行例如视觉感知、语言理解和决策制定等任务(Jordan & Mitchell, 2020)。AI 可以分为不同类型，包括实体 AI(如机器人)，虚拟 AI(如 Siri 和其他数字助手)，以及嵌入式 AI(如汽车导航系统)(Glikson & Woolley, 2020)。在此基础上，文化心理学专注于研究文化如何塑造和影响人类的心理过程，探索这些不同类型的 AI 如何在不同文化背景下被理解和应用(Vygotsky, 1978)。在这一学科中，文化不仅被视为影响个体行为和思维的外部因素，而且被认为是心理过程自身的内在组成部分。鉴于当今 AI 的研究和应用越来越多地触及到了人的心理和社会层面，从文化心理学视角审视 AI 变得尤为重要。

文化心理学能帮助我们理解在全球化的背景下，AI 技术如何与不同文化背景下的用户相互作用，并指导 AI 技术的发展以适应这些交互(Vasalou et al., 2010)。通过考虑文化在塑造人类认知和行为中的作用，AI 的开发可以更好地适应全球用户的多样性需求，从而促进技术的普适性和接受度。文化历史活动理论(CHAT)，亦称为社会文化理论，提出了一个基本假设：人的认知和发展是在社会互动和文化工具的使用中，在特定的文化背景下形成和发展的(Yrjö Engeström, 2019)。这一理论框架强调了人类心理过程与其所处文化环境之间的密切联系，揭示了个体认知发展不可分割的社会文化基础(Cole & Engeström, 1993)。然而，当前 AI 系统的设计和开发常常依赖于大数据和机器学习算法，这种方法在处理复杂的人类行为时可能忽视了文化差异带来的细微变化。这种忽视可能导致 AI 系统无法充分理解和适应不同文化背景下的用户需求，进而加剧文化偏见，影响系统的公正性和有效性(Bhalla et al., 2021)。在大型语言模型（LLMs）的训练和应用中尤其如此。近期研究表明，这些模型的培训数据集主要来源于特定文化背景，特别是美国，从而在决策和语言输出中引入了文化偏见。2023 年的两项重要

研究，来自哥本哈根大学的霍夫斯泰德文化调查和人工智能初创公司 Anthropic 使用的世界价值观调查，对此问题进行了量化分析。哥本哈根大学的研究者通过霍夫斯泰德文化调查对 LLMs 进行了测试，该调查衡量不同国家的人类价值观，结果显示 LLMs 与美国文化有着强烈的一致性（Cao et al., 2023）；Anthropic 公司也利用世界价值观调查进行了类似的测试，得到了相似的结论，即 LLMs 倾向于反映和强化美国文化的各种方面（Anthropic, 2023）。这种文化偏见的存在不仅限于语言模型的语言输出，还影响到模型在解决问题和做决策时的方式。例如，当被要求生成“早餐”的图像时，训练主要在西方图像上的 DALL-E 3 生成了煎饼、培根和鸡蛋的图像，这些选择明显倾向于西方的饮食习惯。这些研究结果凸显了在人工智能技术的开发和应用中，对文化多样性的需求。而未能解决这些偏见可能会在全球应用中造成误解，降低 AI 系统的普适性和效率。

与此同时，人们对 AI 的态度、看法和行为也受到其文化背景的影响。例如，某些文化可能更加重视隐私保护，对于 AI 在个人生活中的渗透持保留态度，而其他文化可能更加开放，乐于接受 A I 技术带来的便利。这些文化差异反过来又对 AI 的接受度、设计要求和应用场景产生了影响，要求开发者在设计 AI 系统时充分考虑多元文化背景下人们的需求和期望(Kim et al., 2022)。因此，要实现真正公正和有效的 AI 系统，既需要 AI 技术能够适应和理解不同文化背景下的需求，也需要深入了解和尊重不同文化背景中人们对 AI 的态度和行为差异。

在实际应用中，将共享和差异的文化部分考虑在内对于 AI 应用的成功至关重要。这不仅有助于完善 AI 的跨文化管理，还能帮助 AI 技术的设计者通过深入了解和整合不同文化中的行为模式和心理特征，重新审视 AI 相关的道德伦理问题，推动人与 AI 的和谐共存。目前的研究发现文化差异可能导致不同文化群体在对待人工智能时持有不同的伦理规范和价值取向，比如不同文化用户对于 AI 隐私和信息分享的期望存在显著差异(Eslami et al., 2015; Vannucc et al., 2019)。因此，从文化心理学的视角审视 AI 不仅有助于保证 AI 系统的公正性和有效性，还是确保 AI 技术能够被广泛接受并在全球范围内的不同文化环境中有效应用的关键。只有通过充分考虑并整合文化多样性，AI 技术才能实现其真正的全球化潜力，并在尊重所有用户的文化背景的同时，提供高质量和包容性的服务。

尽管目前针对各个领域人工智能的开发已有大量研究和发现，其文化层面的影响和应用仍未受到足够的研究关注。尤其是在全球化背景下，AI 如何适应和反映不同文化的共性与差异，仍是一个值得探索的研究方向。当前文献在区分和详细研究文化差异与文化共享如何影响人们对 AI 的态度和应用方向方面存在显著缺口。理论研究和实际应用均未充分重视从文化心理学视角出发来设计和评估 AI 技术的必要性。鉴于此，本文旨在基于现有的研究进行文献评述，为人工智能的跨文化伦理发展，监督管理和全球化应用提供发展意见。本文将先整合社会心理学理论，考察不同文化背景下独特的心理和行为特征，以及这些特征如何影响 AI 技术的监管和应用研究，再研究不同文化下相同的普世价值观对 AI 共有的态度和伦理要求，探讨它们在 AI 技术接受度和伦理原则发展的角色。最后，本文将提出未来研究方向，强调对不同文化特征的考量，以及探索样本多样性对于 AI 研究的影响。

2 人类共享

行为与动机作为心理学的重要研究领域，很多心理学家根据自己的研究提出了不同的观点。例如自我决定理论(SDT)认为人类存在内部和外部两种动机(Deci & Ryan, 1985)以及认为人类存在生存、相互关系、成长三种核心需要的 ERG 理论(Alderfer, 1969)。这些理论都对人类动机与行为之间的关系进行了假设与解释，但是由于研究与实验的样本单一等因素，这些被广泛认可的理论都存在跨文化适用性的局限性。因此越来越多的研究发现，文化是一个对行为及动机有重要影响的因素(Heine, 2007; Steers & Sánchez-Runde, 2017)。Barling(1981)研究发现，需求层次理论在南非不同文化背景群体中，生理、安全和保障、爱和归属感需求预测了黑人群体的行为，而只有生理需求预测了印第安人受试者的晋升愿望。另一方面，白人群体的任何需求不具备预测效果。另一项研究也发现，在集体主义文化中，自尊不属于基本需求，归属感也被发现时基本需求之一(Gambrel & Cianci, 2003)。这些研究都发现，文化价值观影响动机和行为，因此可能对人们与人工智能的预期和应用造成影响。但是在不同文化中，会存在“共享“的价值观。Schwartz(1992)提出了 Theory of basic human values，理论中包含 10 种普世价值观(Universal values)并在 2012 年将其完善至 19 种普世价值观(Schwartz et al., 2012)。与对于需求层次理论的跨文化研究类似，不同文化对这些普世价值观的偏好或重要性存在差异，但不同文化下的人群会在普世价值观的影响下产生相同的动机与行为(Agras & Ates, 2015)，并且在人工智能的开发和应用领域也有相同

的现象存在。全球各地政府及组织对于人工智能存在相同期望与应用需求，因此在人工智能伦理原则与指导中也存在相同的部分，例如 Joblin 等人(2019)发现，在研究涉及到的 84 个来源于各个国家及国际组织的指导纲领中，对透明度的要求出现了 73 次、正义和公平的需求出现了 68 次、对于非恶意和责任的要求均出现了 60 次。在我国近期颁布的《人工智能北京共识》与《新一代人工智能伦理规范》也都规定了人工智能在信息透明，公平以及责任等方面的发展意见。不仅是在规范与指导的制定方面，不同文化也在普世价值观的影响下，对于 AI 的应用也产生了相同的需求，例如在环境、医疗等领域，不同文化的人工智能开发者都期望人工智能在环境治理及医疗辅助能带来积极影响。

2.1 文化共享的价值观

价值观作为一个抽象的概念，被认为是人类决定和调整行为，以及评估自己，他人，外在事物的重要参考。文化心理学研究者认为，价值观作为文化最重要一部分，对该文化下的人群有不可或缺的影响，会潜移默化的影响着某文化环境中群体的习惯，行为倾向以及态度(Russo et al.,2022)。为了探究不同文化下的群体是否有共享的价值观以及这些潜在的共享价值观与人类行为之间的联系，Shalom Schwartz（1992）开发了基本人类价值理论。这是一种跨文化心理学视角下的研究普遍价值观的理论。这个理论扩展了之前的跨文化交流框架，Schwartz 的基本人类价值理论提供了一个普遍模型，概述了指导人类行为的十大广泛价值观，包括权力、成就、享乐主义、刺激、自我导向、普遍主义、仁爱、传统、一致性和安全。这 10 个动机上不同的、广泛的和基本的价值观源于人类条件的三个普遍要求：作为生物有机体的个体需求、协调社会互动的要求、以及群体的生存和福利需求。人类普遍价值观理论关注所有文化中的个体都能认识到的基本价值观。在 Schwartz（2012）的普世价值观里，安全（Security）这个价值观源于个体和群体的基本需求，这些需求包括对物理安全的追求，如避免危险和威胁，以及对心理安全的追求，如稳定的关系和可预测的环境。在新冠疫情期间，AI 机器人的清洁和消毒功能，在物理上帮助高接触高人际的行业避免病毒的威胁，例如酒店(Gursoy & Chi, 2020)、机场(Zeng et al., 2020)，也同时满足了人们的心理安全的追求。除此之外，英国的医院药房利用 AI 机器人辅助配药，成功减少了配药错误率(Goundrey-Smith, 2019)，保护了患者的身心安全。安全价值观体现了不同文化下

人们对社会秩序和保护机制的期望。这一价值观的核心是追求社会、关系和自我的安全、和谐和稳定，强调对稳定和有序环境的需求，并关注保护个人和他人免受伤害。

研究者通过检验在不同的文化中该理论的有效性，发现了普遍价值观适用于不同文化下的不同场景(Datler et al., 2013;Vassallo & Saba,2015)。根据这些普遍价值观，不同文化下的人群产生了相同的行为。Primc et al.(2021)通过跨文化的大样本数据发现，人类基本价值观理论在预测人类的亲环境行为的功能上存在较好的跨文化适用性。在各种企业工作环境中，该理论也发挥了重要的作用。Agras 和 Ates(2015)提出并证实普遍价值观与工作相关行为例如敬业度，组织公民行为，反生产力等行为的关系。Ariza-Montes et al.(2017)也发现了基本价值观与酒店员工责任感及权力使用的关系。

与此同时，普遍主义（Universalism）价值观(Schwartz，2012)作为普世价值观之一，它在推动各国利用 AI 辅助环境治理方面发挥着重要作用。普遍主义价值观的核心目标是理解、欣赏、容忍和保护所有人和自然的福祉。这种价值观超越了对特定群体或社区的关注，扩展到对全人类和环境的广泛关怀。普遍主义价值观源自人类对生存和资源稀缺性的认识，以及对不同个体和群体之间相互依赖的理解。据 Nishant 等人(2020)的研究，人工智能(AI)带来的创新解决方案正在为商业实践和社会问题提供新的方向，为各国追求可持续发展提供支持。Ahmad 等人(2021)发现，AI 能够通过自动化控制能源供需，促进决策和运营流程，改变消费习惯，以实现环境的可持续性，同时最大化运营效率和绩效。

普遍主义价值观强调的是一种全球视角，它超越了个人和小群体的利益，强调全人类的共同利益和对地球资源的共同责任。这种价值观包含了对正义和责任的承诺，意味着在利用 AI 技术时，我们不仅要关注技术带来的直接益处，还要考虑其对环境和社会的长期影响。正义要求我们在利用 AI 时确保公平和平等，保护弱势群体的权益，防止技术滥用和不公平的资源分配。责任则要求我们在开发和部署 AI 解决方案时，考虑到其对环境的影响，确保技术的可持续性和道德性。因此，普遍主义价值观在推动 AI 辅助环境治理的应用中，不仅体现了对技术进步的追求，还体现了对全球正义和责任的重视。这种价值观鼓励我们在追求技术创新的同时，不忘对全人类和自然环境的关怀和保护，确保科技发展与社会福祉和环境可持续性相协调，创造一个更加公正、安全和可持续的未来。

2.2 跨文化的 AI 伦理共识

在 AI 领域，伦理因为与普遍主义价值观紧密相连而成为共同的重要概念。普遍主义价值观强调的是整体福祉(Schwartz，2012) 。在全球范围内，AI 的伦理指导和政策制定显现出了文化共享的价值观。尽管不同文化背景下的人们对于 AI 技术有着不同的接受度和适应方式，全球社会对于 AI 的透明度、公平、正义和责任等方面表现出了显著的一致性(ÓhÉigeartaigh et al., 2020)。这种普遍的共识反映了人类对于 AI 技术应如何被开发和应用的基本价值观。其中，透明度要求 AI 的决策过程能够被理解和审查，这对于建立用户对 AI 系统的信任至关重要。例如，Jobin 等人(2019)的研究显示，在全球范围内的 AI 伦理指导原则中，透明度是被最多提及的原则之一。透明度不仅有助于揭示 AI 决策的基础和逻辑，也使得潜在的偏见和错误能够被及时识别和纠正，从而增强了 AI 系统的公信力和接受度；正义和公平则强调 AI 决策不应加剧现有的不平等，而应努力减轻社会不公；责任的要求确保在 AI 系统出现错误或不当行为时，可以明确归责并采取适当的纠正措施，这不仅包括技术层面的错误修正，也涉及对受影响个体和群体的补偿和保护。由此可见，全球范围内对于 AI 的透明度、公平、正义和责任的共同要求，反映了一种跨文化的、普遍接受的对 AI 技术应有的伦理标准。

在这种伦理框架下，Ikkatai 等人（2022）的研究进一步揭示了人类的普遍价值观如何在 AI 的具体应用和伦理指导原则中体现。他们聚焦于 AI 指导原则中普遍共享的八个主题：隐私、问责制、安全与保障、透明度与可解释性、公平与非歧视、技术的人类控制、专业责任和人类价值的提升。研究通过在日本进行的四种场景下的在线问卷调查，探讨了公众对 AI 伦理的态度，发现公众对于使用 AI 的赞同或反对程度因场景而异。例如，在 AI 用于武器系统的场景中，人们对 AI 伦理的担忧程度较高；年龄因素显著影响了人们在不同场景中对这些主题的看法，而性别和对 AI 的理解则因主题和场景而有所不同（Ikkatai et al., 2022）。我们不仅看到了人类普遍价值观如安全、正义和责任在 AI 政策制定中的反映，也观察到这些价值观与 AI 伦理原则之间的交叉和重合。

通过分析和理解不同文化对 AI 伦理原则共通的看法，我们可以更深入地探讨如何在全球范围内推广这些原则，以确保 AI 技术的伦理发展得到广泛的文化支持和认可。这种跨文化的伦理共识也为 AI 的全球应用和政策制定提供了坚实的道德基础，有助于推动 AI 技术进步与伦理考量在不同文化中的同步发展。

3 文化差异

文化差异理论涉及多种框架和模型，旨在研究和阐明不同文化之间的多样性和差异。其中，荷兰社会心理学家 Geert Hofstede 提出的“文化维度理论” (Geert Hofstede's Cultural Dimensions Theory)具有深远的影响，它将文化划分为六个可进行比较的维度，分别是权力距离，不确定性回避，个体主义与集体主义，男性化与女性化，长期导向与短期导向，满足与约束(Hofstede, 2011)。

不同文化在每个维度上存在差异。例如，在权力距离方面，东方文化更倾向于高权力距离，更加接受权力不平等，而西方文化更倾向于低权力距离。在不确定性回避方面，拉丁文化、日本和德国文化更倾向于高不确定性回避，强调规则和稳定，而中国和北欧文化更倾向于低不确定性回避。在个体主义与集体主义方面，西方文化更倾向于个体主义，强调个人利益，而东方文化更倾向于集体主义，强调集体利益。在男性化与女性化方面，日本、德国等文化更倾向于男性化，即男性占主导地位，而北欧、荷兰等文化更倾向于女性化。在长期导向与短期导向方面，东亚文化更倾向于长期导向，更注重个人努力、积蓄，而美国、澳大利亚等文化更倾向于短期导向，注重消费主义和个人运气。在满足与约束方面，北美、西欧文化更倾向于满足，强调自由开放，而东欧、亚洲、伊斯兰文化更倾向于约束，强调自律和规范。

通过梳理以往研究，我们发现人对人工智能的态度确实受到文化差异的影响，且有学者使用 Hofstede 的文化维度理论并通过实证研究证明了各个维度不同的文化特点影响了人对人工智能的态度(Chi et al., 2023)以及人和人工智能系统的互动 (Lee & Joshi, 2020)。Chi 等人(2023)发现不确定性规避、长期导向和权力距离这三个文化维度在酒店顾客使用 AI 机器人意愿方面发挥了显著的作用，而 Lee 和 Joshi(2020)发现不确定性回避、个体主义或集体主义两个维度影响了用户与人工智能系统的互动。其他研究虽没有直接使用 Hofstede 的文化维度理论阐明人工智能方面的文化差异，但通过直接比较不同文化的被试以及通过总结前人文献，也发现了人工智能在文化差异方面的表现规律。

3.1 东西方文化在对待机器人和人工智能方面的态度差异

现有的大部分文献都表明东方人比西方人更接受人工智能(Sindermann et al., 2022; Yam et al., 2023)。研究显示，中国人对于人工智能的接受度高于德国人和英国人，而

恐惧度则低于德国人和英国人。Yam 等(2023)通过总结现有的研究发现，东方文化更倾向于将机器人视为自然界的一部分，更接受人工智能和机器人，而西方文化更倾向于将它们视为外来者。他们提出了一个东西方的历史、宗教和文化曝光度理论框架来解释东西方对人工智能态度的不同(Yam et al., 2023)。历史框架指的是东方的万物有灵论传统和西方的人本主义传统分别影响了两种文化对机器人的态度。宗教框架指的是东方的佛教和神道教与西方的基督教分别强调了不同的人与非人类实体的思想关系，这影响了东西方文化对待机器人不同的态度。例如，在日本，人们经常认为非人类实体也具有灵魂或精神，这源于神道教的影响。相反，西方文化更倾向于将机器人和人工智能视为外来者，这与基督教强调人的独特性和重要性有关。曝光度框架指的是，东方人接触机器人的机会更多，这也有助于减少对机器人的厌恶感，比如日本的机器人产业由来已久，这也与这与 Sindermann 等人(2022)对东方人比西方人更接受人工智能这一现象推测的原因一致。总的来说，我们发现东方文化相比西方文化对人工智能接受度更高，是因为东方文化国家如日本和中国在日常生活中用人工智能更多，且东方文化的宗教和历史将非人类实体视为自然界的一部分，甚至赋予非人类实体如人工智能思想和灵性。和西方文化不同，东方文化不将人和非人类实体如人工智能严格区分开来。与西方文化更认可人作为人的独特性不同，东方文化倾向认为万物皆有灵，因而也更接受人工智能的存在而不将其视为威胁和他者。

除了对人工智能接受度的差异，其他研究还从态度的其他方面如预期和情感态度，来表明东西方文化在对待人工智能上的差异。Haring 等人(2014)的研究比较了日本人和欧洲人对机器人的预期和态度，揭示了两者在多个方面的显著差异。日本人倾向于将机器人视为机器而非人形，而欧洲人没有明显的偏好。和 Yam 等(2023)的“万物有灵论”理论框架推理的不同，Haring 等人(2014)发现日本被试并没有倾向于将机器人视为人形，而是倾向于将机器人视为机器而非人形，而欧洲人没有明显的偏好。在负面看法上，Haring 等人(2014)发现日本被试并没有不担心机器人的滥用问题，也没有如欧洲人对机器人发展持乐观态度，而是对机器人的普及和发展持更谨慎的态度。此外，在预期任务上，日本人更期望机器人能提供按摩等亲密服务，而欧洲人对此没有特别的需求。在交互方式上，日本人更喜欢机器人能够自主完成任务，而欧洲人则更倾向于通过语音与机器人交互。正面看法方面，两者都认为机器人能够帮助完成日常琐事，但日本人特别强调机器人在医疗和辅助残疾人方面的作用。在负面看法上，日本

人更担心机器人的滥用问题，而欧洲人则更担心机器人会取代人类的工作。对未来的预期显示，日本人对机器人的普及和发展持更谨慎的态度，而欧洲人则相对乐观。在外观设计上，日本人和欧洲人都喜欢机器人的外观设计更接近机器而非人形。在对人工智能的情感态度上，Kim 等 (2022)的研究发现韩文用户在人工智能方面更看重愉悦性，而英文用户更看重实用性。同时，英文用户比韩文用户更看重风险维度。

3.2 用户与人工智能系统的互动方面存在文化差异

文化背景影响了用户与人工智能系统(AIS)的互动，因为文化价值观影响了用户使用 AIS 的决策(Lee & Joshi, 2020)。他们指出不确定性回避程度高的文化背景下的用户更可能依赖 AIS，而个体主义文化背景下的用户则更倾向于自主决策。这一发现与 Hofstede 的文化维度理论相呼应，该理论认为不同文化背景下的个体在面对不确定性时的行为倾向存在差异(Hofstede， 2011)。同时，集体主义文化的用户可能更注重社会和谐和群体利益，因此在 AIS 的推荐中可能更偏好那些能够促进社会联系和群体福祉的选项。在使用模式方面，集体主义文化的用户在使用 AIS 时可能会表现出不同的动态模式。这些用户可能更能够处理相互矛盾的信息，并在决策中考虑多种可能性。相反，个体主义文化的用户可能更倾向于在两个相反的陈述中选择一个，并排除另一个，以减少认知失调。因此，当 AIS 提供的推荐确认了他们的预期决策时，个体主义文化的用户可能会更频繁地使用 AIS。上述研究发现说明，不同文化在文化维度如不确定回避程度和个体主义、集体主主义的区别，会导致不同文化的用户与人工智能互动层面（使用人工智能进行决策、对人工智能的依赖度、在人工智能系统推荐中的偏好、使用模式与频率）的差异。

3.3 人工智能政策制定存在文化差异

在人工智能政策制定，如人工智能的伦理和治理，东西方也存在差异(ÓhÉigeartaigh et al., 2020; Wong, 2020)。这种文化差异导致了在国际人工智能伦理和治理合作上的挑战，尤其是如何在制定全球标准和保留不同文化需求之间平衡(ÓhÉigeartaigh et al., 2020; Wong, 2020)。Wong(2020)指出，文化差异可能使一些行为者用来为违反重要伦理价值的行为忽视或辩解，这对全球 AI 的伦理和治理构成了严重挑战。例如，一些文化可能缺乏特定的伦理价值(如隐私)，或者持有与西方价值观相冲突的价值观(如认为国家干预是好的)。论文强调，尽管人权方法旨在提供一个普遍适用且全球可执行的框

架，但其在处理文化差异的挑战时尚未充分考虑文化多样性的影响，且人权方法可能在非西方文化背景下难以直接应用(Wong, 2020)。因此，全球 AI 技术的伦理和治理的规范标准必须考虑文化多样性，且不应被视为预定的终点，而应被视为一个持续的谈判和构建过程。

现有的研究强调，全球人工智能政策制定者需要对文化价值保持开放和响应，并与其他文化合作，以丰富全球人工智能伦理和治理的规范基础(Wong，2020)。为确保全球人工智能政策的制定能够真正反映和尊重全球文化多样性，ÓhÉigeartaigh 等人(2020)的研究分析了欧洲、北美与东亚之间在人工智能伦理和治理合作方面的障碍，并提出了一系列实际建议来促进跨文化合作，包括关键文件的多语言翻译、研究者交流项目和跨文化研究议程的发展。他们认为，尽管存在误解和文化差异，但可以通过学术界的努力建立更大的相互理解和信任，从而推动有效的跨文化合作。他们强调，合作并不要求在所有人工智能领域的伦理原则上达成一致，而是可以在实际问题和应用上寻求共识，例如，尽管在数据隐私等关键问题上存在价值观差异，但不同文化和社会可以在保护个人隐私的共同目标上达成一致。这为国际合作提供了可行的路径，并为我们提供了如何避免文化差异影响全球人工智能政策制定的重要参考。

4 贡献与启示

本文从文化角度切入，探讨了文化共享价值观如何影响和引导人工智能的应用需求以及伦理准则的制定；同时，通过概述不同文化背景下人们对人工智能的态度、互动行为倾向以及政策需求，揭示了各文化间的差异性。这项工作将有助于推动人工智能心理学领域的研究进展，并且能够帮助我们更好地理解文化环境对人工智能开发和应用的意义，从而推动人工智能发展以及人与人工智能共存。

基于前文对既往研究的总结，我们发现无论处于哪种文化背景，安全与普遍主义价值观都广泛存在，这些价值观指导着人工智能在疫情期间服务行业和环境可持续方面的应用。此外，人们在隐私、透明度、公平、正义和责任等方面对人工智能有着共同需求和期望，这进一步促进了人工智能伦理标准共识的形成。然而，在实际应用和规范人工智能的过程中，由于历史、宗教和文化曝光度的差异，不同文化背景中的人们仍然表现出了不同的态度、行为以及政策制定倾向。具体而言，东方文化更倾向于接

受人工智能与人类共生共存，对其发展和普及速度持保守态度；而西方文化则更倾向于将人工智能视作与人类相对立，强调其对人类群体构成的威胁，认为人工智能的未来发展将进展迅速。在与人工智能互动过程中，集体主义文化背景下的个体更依赖人工智能的决策，能够同时思考相互矛盾的信息；而个体主义文化背景下的个体更倾向于自主决策，并且偏好在矛盾信息中选择一个方向。此外，东西方文化在人工智能政策制定时遵循的原则存在一定差异，西方文化强调个体隐私和数据透明，而东方文化则更注重社会稳定与国家安全，支持国家干预。在此基础上，本文还对未来人工智能心理学研究方向提出了一些思考和建议。

5 局限与未来展望

首先，现有的文献大多直接采用东方国家和西方国家的二元分类，探索东西方文化的共享与差异，但现有的研究较少对文化进行更为细致的量化测量。由于全球化的影响，东西方差异可能也在逐步弥合。直接采用既定的分类方法，可能无法完全描述研究国家所处的文化(Kirkman et al.，2006)。未来的研究应该增加对文化的测量，以进一步探索不同文化中 AI 运用共享与差异的内容。

其次，目前的研究大多关注特定的文化特征。其中，很大一部分研究集中在探究集体主义国家和个体主义国家对于 AI 使用与应用文化共享与文化，其他的文化特征还未得到足够多的重视。比如，已经有初步研究发现不确定性规避可能会影响不同文化背景下人们对于 AI 工具的态度、使用(Lee & Joshi, 2020)。未来的研究可以探索具体的文化特征，比如说权力距离等文化特征对于 AI 运用、AI 影响等所起的作用。

再者，现有的研究样本存在文化偏见。从研究开展的国家来看，目前探究文化共享或者文化差异的研究主要是美国、日本、中国、韩国等国家开展。尽管这些国家是东方和西方国家中比较典型的国家。但是东方国家内或者西方国家内各自之间的差异也很大(Lim et al.，2021)。未来的研究可以进一步在东西方文化下的其他国家开展研究，进一步探索现有的结论是否能够泛化。

最后，目前关于文化在 AI 现象中的影响的研究未能充分考虑其他相关因素。有研究提出了除了文化价值观外，年龄、性别种族、受教育水平等都会对人类与 AI 的交互产生影响 (Mantello et al., 2023; O’Shaughnessy et al., 2022)。比如，有研究表明受到良

好教育、收入较高的群体往往对 AI 工具有着更清晰完整的认识，能够更好地利用 AI 工具，并且受到 AI 的负面影响较小(Mantello et al., 2023)。为了进一步验证目前研究得出的关于文化共享以及文化差异的结论，在未来的研究中需要控制这些潜在的变量，以厘清文化在 AI 现象中的作用。

综上所述，本论文通过对人工智能心理学中的文化共享和文化差异进行了全面探讨，凸显了文化背景在人工智能技术接受度、应用范围以及政策制定中的核心作用。研究指出，虽然不同文化背景的群体在对人工智能的透明度、公正性、责任感等方面有共同的期望，但在人工智能的接受度、人机互动方式以及政策导向上存在明显的文化差异。东方文化倾向于接纳与人工智能的和谐共存，而西方文化对于人工智能可能构成的威胁持更为谨慎的态度。此外，文化背景影响了用户与人工智能系统的互动方式，集体主义文化的个体更依赖人工智能做出的决策，而个体主义文化的个体则更倾向于自主决策。未来研究应进一步探索文化多样性在人工智能设计、开发和政策制定中的作用，以确保人工智能技术能够在全球范围内被公平、有效地应用。同时，需要开发和验证跨文化有效的人工智能应用，以及考虑不同文化背景下的用户需求和预期。通过跨文化合作和交流，可以促进人工智能技术的全球化发展，为不同文化背景下的用户提供更加个性化、包容性强的服务。最后，研究强调了在全球化和多元文化背景下，理解和尊重文化差异，以及培养跨文化沟通与合作能力的重要性。人工智能技术的发展不仅需要技术创新，还需要深入理解人类社会和文化的多样性，以促进人与人工智能的和谐共存，为构建更加公正、平等、包容的全球社会贡献心理学的力量。

参考文献

Agras, S., & Ates, H. (2015). Investigating the predictive abilities of Schwartz's ten universal values on five work performance outcome behaviors. Social and Basic Sciences Research Review, 3(1), 16-31.

Ahmad, T., Zhang, D., Huang, C., Zhang, H., Dai, N., Song, Y., & Chen, H. (2021). Artificial intelligence in sustainable energy industry: status quo, challenges and opportunities. Journal of Cleaner Production, 289, 125834. https://doi.org/10.1016/j.jclepro.2021.125834

Alderfer, C. P. (1969). An empirical test of a new theory of human needs. Organizational Behavior and Human Performance, 4(2), 142–175. https://doi.org/10.1016/0030-5073(69)90004-X

Anthropic. (2023, March 8). Core views on AI safety: when, why, what, and how. Www.anthropic.com. https://www.anthropic.com/news/core-views-on-ai-safety

Ariza-Montes, A., Arjona-Fuentes, J. M., Han, H., & Law, R. (2017). Employee responsibility and basic human values in the hospitality sector. International Journal of Hospitality Management, 62, 78–87.

Barling, J. (1981). A cross-cultural study of Maslow's motivation theory in industry. South African Journal of Psychology, 11(1), 47–50. https://doi.org/10.1177/008124638101100107

Bhalla, K., Shivakumar, S., & Kumar, T. (2021). Design justice: community-led practices to build the world we need (information policy) by Sasha Costanza-Chock. Design Issues, 37(4), 103–107. https://doi.org/10.1162/desi_r_00661

Cao, Y., Zhou, L., Lee, S., Cabello, L., Chen, M., & Hershcovich, D. (2023). Assessing cross-cultural alignment between ChatGPT and human societies: an empirical study. (arXiv:2303.17466). arXiv. https://doi.org/10.48550/arxiv.2303.17466

Chi, O. H., Chi, C. G., Gursoy, D., & Nunkoo, R. (2023). Customers' acceptance of artificiall y intelligent service robots: the influence of trust and culture. International Journal of Information Management, 70, 102623. https://doi.org/10.1016/j.ijinfomgt.2023.10262

3

Datler, G., Jagodzinski, W., & Schmidt, P. (2013). Two theories on the test bench: internal and external validity of the theories of Ronald Inglehart and Shalom Schwartz. Social Science Research, 42(3), 906–925.

Deci, E. L., & Ryan, R. M. (1985). Intrinsic motivation and self-determination in human behavior. Springer Science, LLC. https://doi.org/10.1007/978-1-4899-2271-7

Gambrel, P. A., & Cianci, R. (2003). Maslow's hierarchy of needs: does it apply in a collectivist culture. Journal of Applied Management and Entrepreneurship, 8, 143-161.

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: review of empirical research. Academy of Management Annals, 14(2). https://doi.org/10.5465/annals.2018.0057

Goundrey-Smith, S. J. (2019). Technologies that transform digital solutions for optimising medicines use in the NHS. BMJ Health & Care Informatics, 26(1), e100016. https://doi.org/10.1136/bmjhci-2019-100016

Gursoy, D., & Chi, C. G. (2020). Effects of COVID-19 pandemic on hospitality industry: review of the current situations and a research agenda. Journal of Hospitality Marketing and Management, 29(5), 527–529. https://doi.org/10.1080/19368623.2020.1788231

Haring, K. S., Mougenot, C., Ono, F., & Watanabe, K. (2014). Cultural differences in perception and attitude towards robots. International Journal of Affective Engineering, 13(3), 149-157.

Heine, S. J. (2007). Culture and motivation. Handbook of cultural psychology, 714-733.

Hofstede, G. (2011). Dimensionalizing cultures: the Hofstede model in context. Online readings in psychology and culture, 2(1), 8.

Ikkatai, Y., Hartwig, T., Takanashi, N., & Yokoyama, H. M. (2022). Octagon measurement: public attitudes toward AI ethics. International Journal of Human–Computer Interaction, 38(17), 1–18. https://doi.org/10.1080/10447318.2021.2009669

Jobin, A., Ienca, M., & Vayena, E. (2019). Artificial Intelligence:the global landscape of ethi cs guidelines. (arXiv:1906.11668). arXiv. https://doi.org/10.48550/arxiv.1906.11668

Jordan, M. I., & Mitchell, T. M. (2020). Machine learning: trends, perspectives, and prospects. Science, 349(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Kim, J. H., Jung, H. S., Park, M. H., Lee, S. H., Lee, H., Kim, Y., & Nan, D. (2022). Explorin g cultural differences of public perception of artificial intelligence via big data approach. International conference on human-computer interaction, 427-432. Springer International Publishing.

Kirkman, B. L., Lowe, K. B., & Gibson, C. B. (2006). A quarter century of culture's consequ ences:
a review of empirical research incorporating Hofstede's cultural values framework. Jo urnal of International Business Studies, 37, 285-320.
https://doi.org/10.1057/palgrave.jibs.8400202

Lee, K., & Joshi, K. (2020). Understanding the role of cultural context and user interaction in artificial intelligence-based systems. Journal of Global Information Technology Management, 23(3), 171-175.

Lim, V., Rooksby, M., & Cross, E. S. (2021). Social robots on a global stage: establishing a r ole for culture during human–
robot interaction. International Journal of Social Robotics, 13(6), 1307–
1333. https://doi.org/10.1007/s12369-020-00710-4

Mantello, P., Ho, M. T., Nguyen, M. H., & Vuong, Q. H. (2023). Bosses without a heart: soci o-
demographic and crosscultural determinants of attitude toward Emotional AI in the w orkplace. AI & Society, 38(1), 97-119.

Maslow, A. H. (1943). A theory of human motivation. Psychological Review, 50(4), 370–396. https://doi.org/10.1037/h0054346

Nishant, R., Kennedy, M., & Corbett, J. (2020). Artificial intelligence for sustainability: challenges, opportunities, and a research agenda. International Journal of Information Management, 53, 102104. https://doi.org/10.1016/j.ijinfomgt.2020.102104

O'Shaughnessy, M. R., Schiff, D. S., Varshney, L. R., Rozell, C. J., & Davenport, M. A. (2023). What governs attitudes toward artificial intelligence adoption and governance? Science and Public Policy, 50(2), 161-176.

ÓhÉigeartaigh, S. S., Whittlestone, J., Liu, Y., Zeng, Y., & Liu, Z. (2020). Overcoming barriers to cross-cultural cooperation in AI ethics and governance. Philosophy & technology, 33, 571-593.

Primc, K., Ogorevc, M., Slabe-Erker, R., Bartolj, T., & Murovec, N. (2021). How does Schwartz's theory of human values affect the proenvironmental behavior model? Baltic Journal of Management, 16(2), 276–297.

Russo, C., Danioni, F., Zagrean, I., & Barni, D. (2022). Changing personal values through value-manipulation tasks: a systematic literature review based on Schwartz's theory of basic human values. European Journal of Investigation in Health, Psychology and Education, 12(7), 692–715.

Schwartz, S. H. (1992). Universals in the content and structure of values: theoretical advances and empirical tests in 20 countries. Advances in Experimental Social Psychology, 25, 1-65. Elsevier Science & Technology. https://doi.org/10.1016/S0065-2601(08)60281-6

Schwartz, S. H., Cieciuch, J., Vecchione, M., Davidov, E., Fischer, R., Beierlein, C., Ramos, A., Verkasalo, M., Lönnqvist, J.-E., Demirutku, K., Dirilen-Gumus, O., & Konty, M. (2012). Refining the theory of basic individual values. Journal of Personality and Social Psychology, 103(4), 663–688. https://doi.org/10.1037/a0029393

Sindermann, C., Yang, H., Elhai, J. D., Yang, S., Quan, L., Li, M., & Montag, C. (2022). Acceptance and fear of artificial intelligence: associations with personality in a German and a Chinese sample. Discover Psychology, 2(1), 8.

Steers, R. M., & Sánchez-Runde, C. J. (2017). Culture, motivation, and work behavior. The Blackwell handbook of cross-cultural management, 190-216. https://doi.org/10.1002/9781405164030.ch10

Vannucci, F., Sciutti, A., Lehman, H., Sandini, G., Nagai, Y., & Rea, F. (2019). Cultural differences in speed adaptation in human-robot interaction tasks. Journal of Behavioral Robotics, 10(1), 256–266. https://doi.org/10.1515/pjbr-2019-0022

Vasalou, A., Joinson, A. N., & Courvoisier, D. (2010). Cultural differences, experience with social networks and the nature of “true commitment” in Facebook. International Journal of Human-Computer Studies, 68(10), 719–728. https://doi.org/10.1016/j.ijhcs.2010.06.002

Vassallo, M., & Saba, A. (2015). Does money for grocery expenditure sway Italian consumers’ motivational values in predicting attitude towards eco-sustainable food products? Contemporary Management Research,11(1), 003–024.

Vygotsky, L. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

Wong, P. H. (2020). Cultural differences as excuses? Human rights and cultural values in global ethics and governance of AI. Philosophy & Technology, 33(4), 705-715.

Yam, K. C., Tan, T., Jackson, J. C., Shariff, A., & Gray, K. (2023). Cultural differences in people's reactions and applications of robots, algorithms, and artificial intelligence. Management and Organization Review, 19(5), 859-875.

Yrjö Engeström. (2019). Learning by expanding: an activity-theoretical approach to developmental research. Cambridge University Press.

Zeng, Z., Chen, P.-J., & Lew, A. A. (2020). From high-touch to high-tech: COVID-19 drives robotics adoption. Tourism Geographies, 22(3), 724–734. https://doi.org/10.1080/14616688.2020.1762118

The Role of Culture in Human-Computer Interaction:

Human Universals and Cultural Differences

Abstract: The rapid advancement and widespread application of Artificial Intelligence (AI) technologies have significantly altered human living and working practices, thereby attracting scholarly attention to its sociocultural implications. This literature review examines the influence of cultural factors on the interpretation, principles, and utilization of AI across diverse contexts, emphasizing the critical interaction between AI technologies and cultural psychological principles. AI encompasses a broad spectrum of capabilities, ranging from basic algorithms to sophisticated machine learning systems, designed for tasks such as sensory perception, linguistic understanding, and decision-making. However, the adoption and integration of these technologies exhibit considerable variation across different cultural environments. The study highlights the importance of integrating cultural perspectives to achieve equitable, effective, and universally acceptable AI systems. Through the lens of cultural psychology, this research provides insights into the development of culturally attuned AI systems. It advocates for a comprehensive understanding of both cultural variances and shared values within the realm of AI. The paper proposes future research directions focusing on incorporating cultural diversity into AI research and applications, aiming to realize the full global potential of AI technologies.

Keywords: artificial intelligence, cultural psychology, cultural differences, cultural sharing, AI ethics